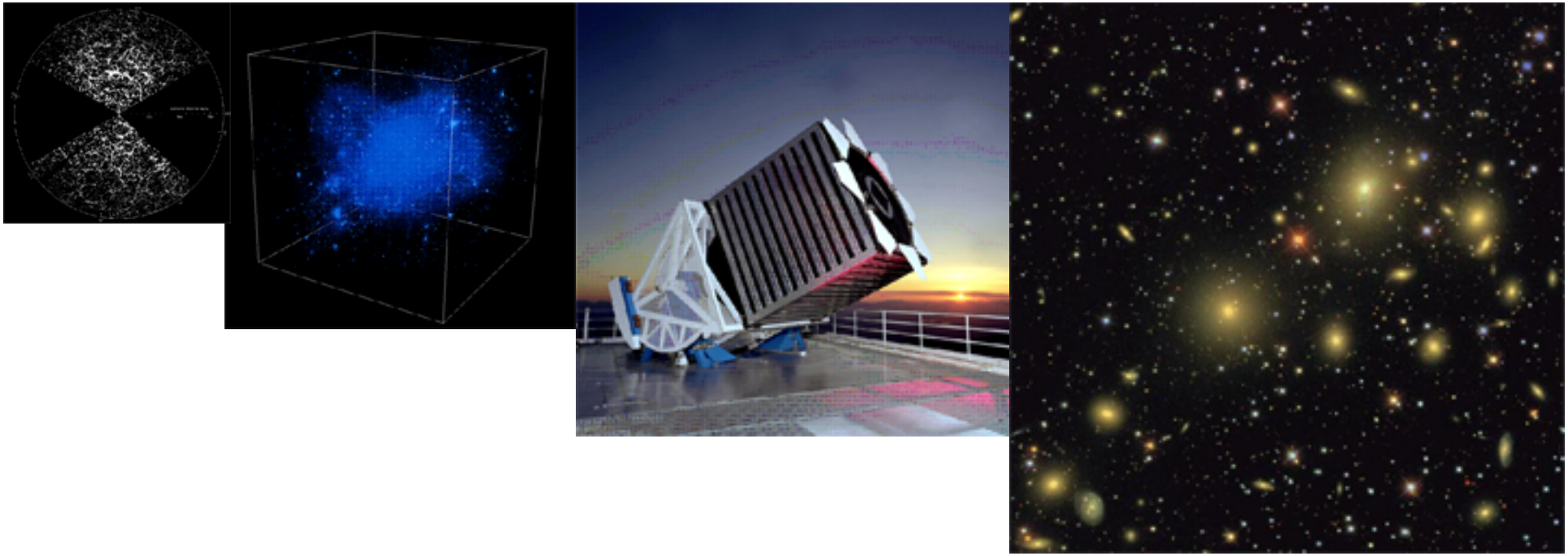


Extreme Data-Intensive Computing in Astrophysics

Alex Szalay
The Johns Hopkins University

The Science of Big Data

- Data growing exponentially, in all science
- Changes the nature of science

  =>from hypothesis-driven to data-driven discovery
- Cuts across all sciences
- Non-incremental!
- Industry and government faces the same challenges
  - *Google, Microsoft, Yahoo, NSA, DOD,…*
  - *Google (~10 Exabytes, many Tbits/s bandwidth)*
- Convergence of physical and life sciences through Big Data (statistics and computing)
- A new scientific revolution

  => a rare and unique opportunity

Non-Incremental Changes

- Science is moving from hypothesis-driven to data-driven discoveries

  Astronomy has always been data-driven….
  now becoming more generally accepted

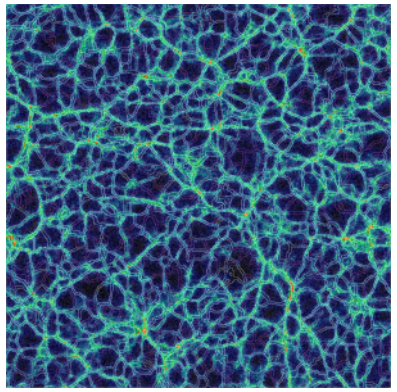
- Need new data intensive scalable architectures
- Need new randomized, incremental algorithms
  - *Best result in 1 min, 1 hour, 1 day, 1 week*
- New computational tools and strategies

  … not just statistics, not just computer science, not just astronomy…

Continuing Growth

How long does the data growth continue?

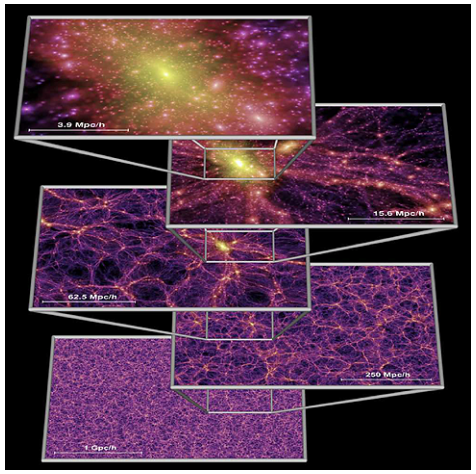
- High end always linear
- Exponential comes from technology + economics
  - *rapidly changing generations*
  - *like CCD's replacing plates, and become ever cheaper*
- How many generations of instruments are left?
- Are there new growth areas emerging?
- **Software is becoming a new kind of instrument**
  - *Value added data*
  - *Hierarchical data replication*
  - ***Large and complex simulations***



Cosmological Simulations

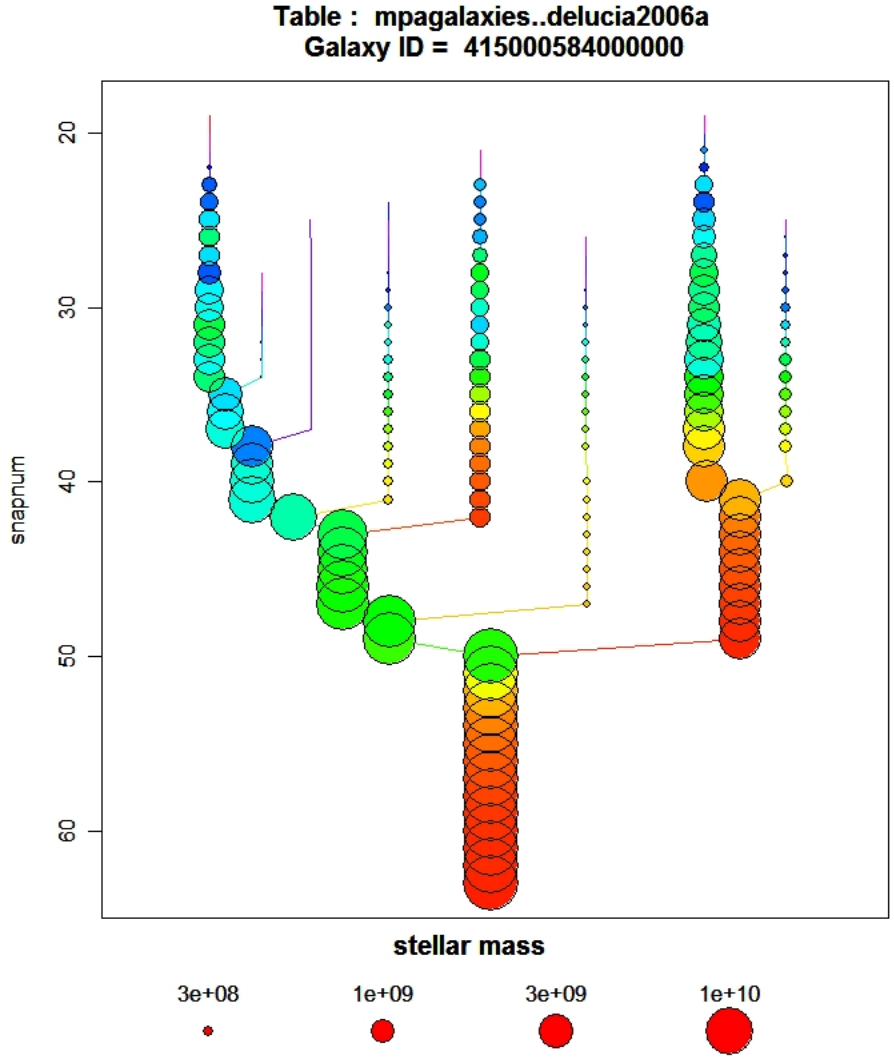
In 2000 cosmological simulations had 10^{10} particles and produced over 30TB of data (Millennium)

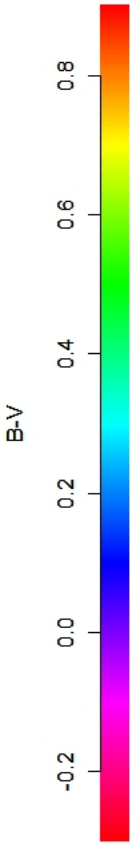
- Build up dark matter halos
- Track merging history of halos
- Use it to assign star formation history
- Combination with spectral synthesis
- Realistic distribution of galaxy types

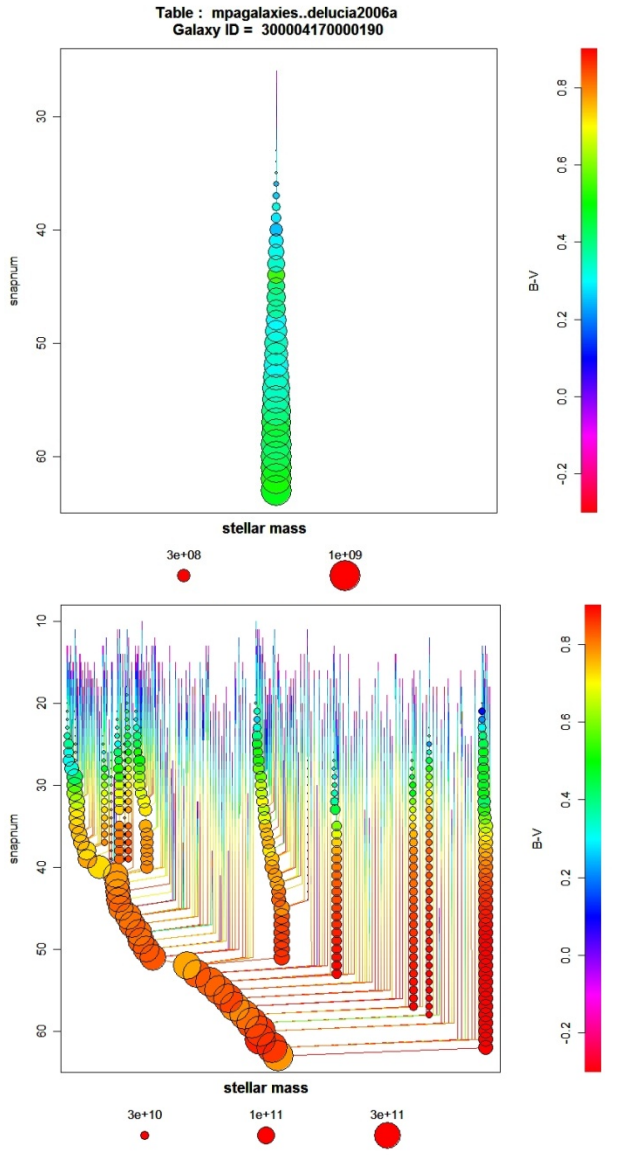


- Today: simulations with 10^{12} particles and PB of output are under way (MillenniumXXL, Silver River, etc)
- Hard to analyze the data afterwards -> need DB
- What is the best way to compare to real data?

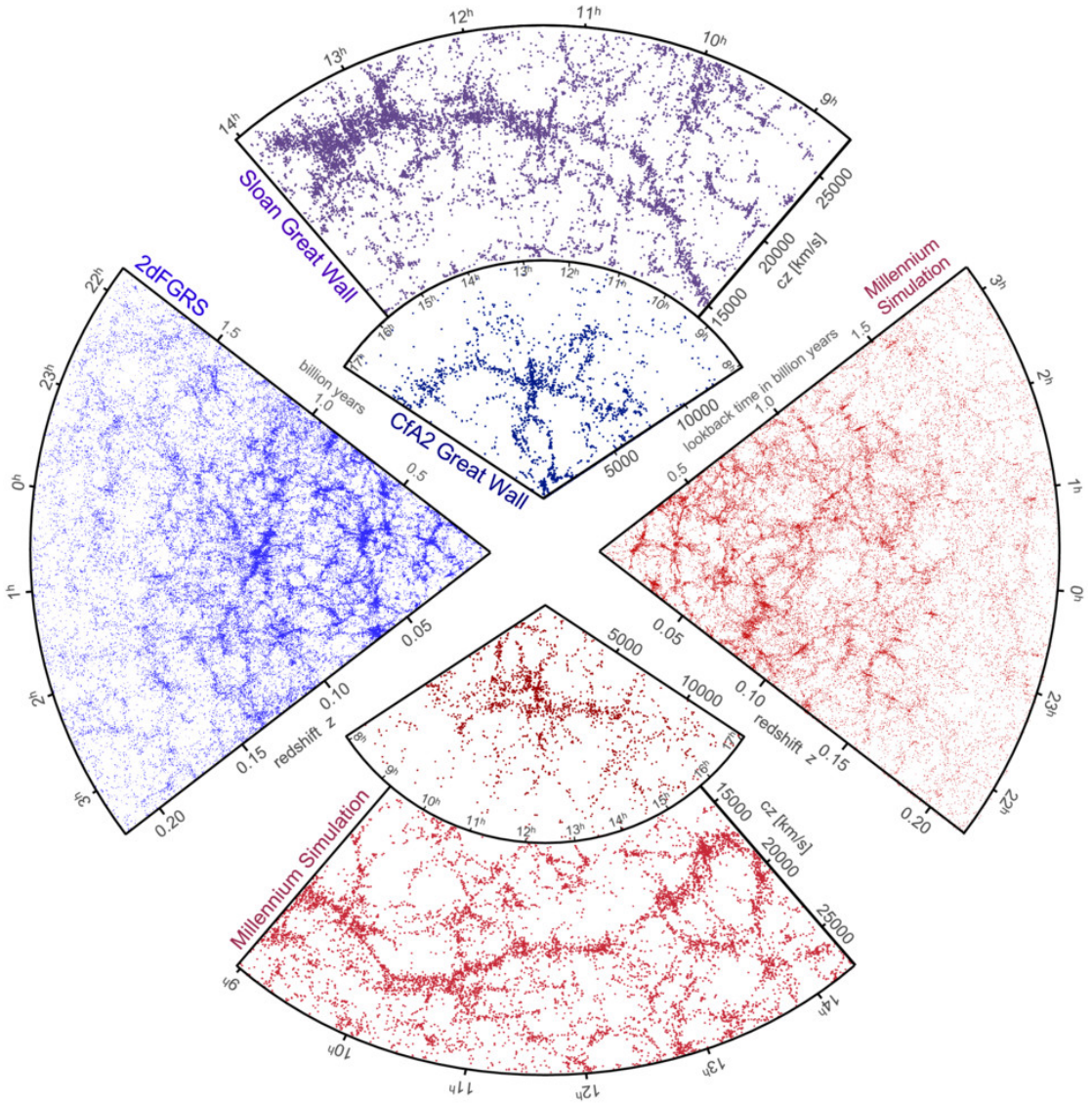
Time evolution: merger trees



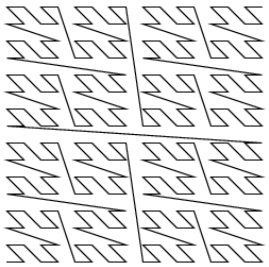




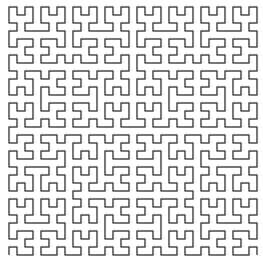
Mock Catalogues



Spatial queries, random samples



- Spatial queries require multi-dimensional indexes.
- (x,y,z) does not work: need discretisation
  - *index on (ix,iy,iz) withix=floor(x/10) etc*
- More sophisticated: space fillilng curves
  - *bit-interleaving/octtree/Z-Index*
  - *Peano-Hilbert curve*
  - *Need custom functions for range queries*
  - *Plug in modular space filling library (Budavari)*



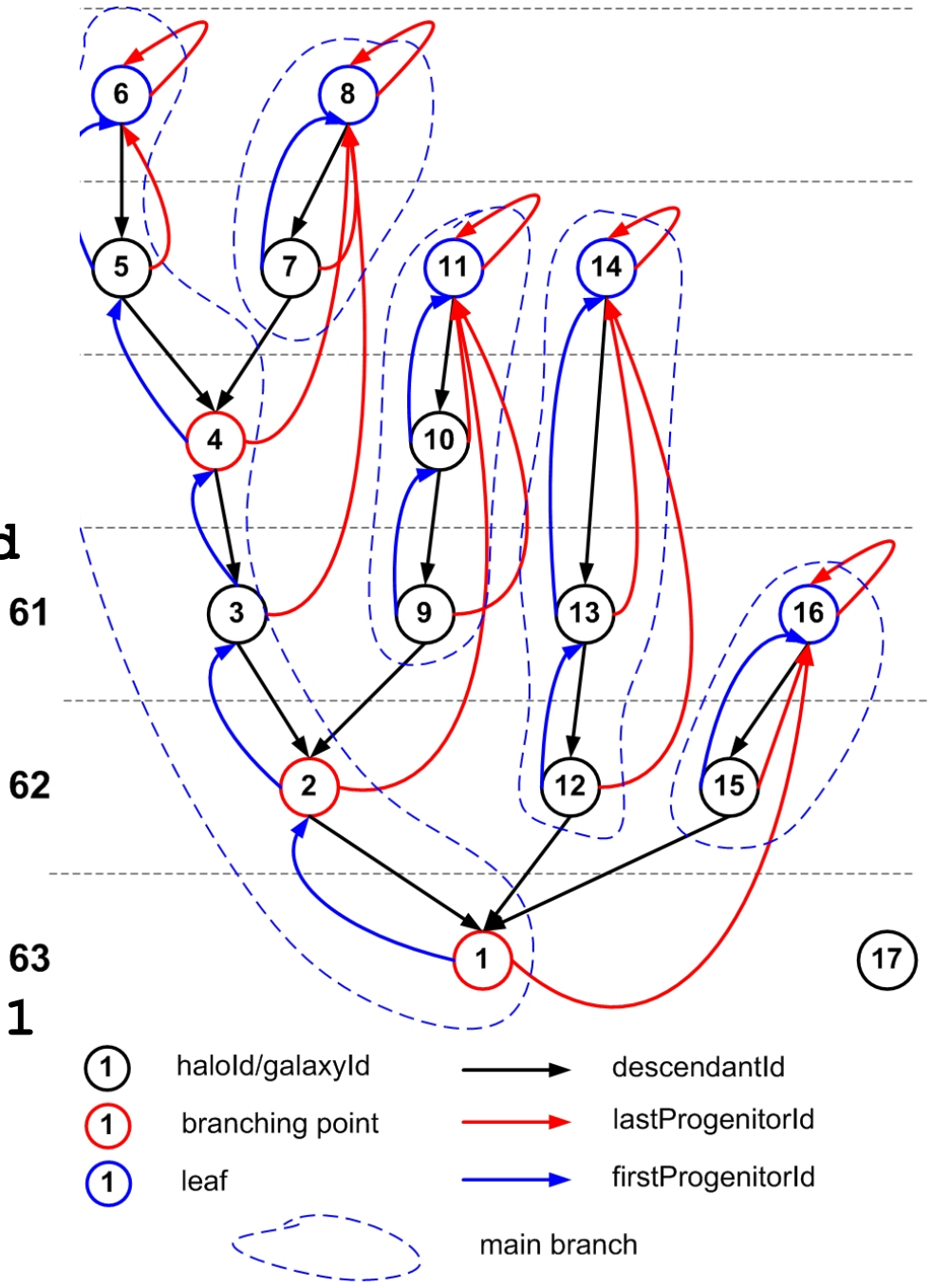
- Random sampling using a RANDOM column
  - *RANDOM from [0,1000000]*

Merger trees :

```
select prog.*
  from galaxies d
  ,    galaxies p
where d.galaxyId = @id
  and p.galaxyId
  between d.galaxyId
  and d.lastProgenitorId
```

Branching points :

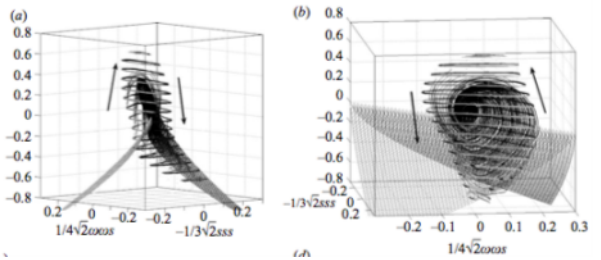
```
select descendantId
  from galaxies d
 where descendantId != -1
 group by descendantId
  having count(*) > 1
```

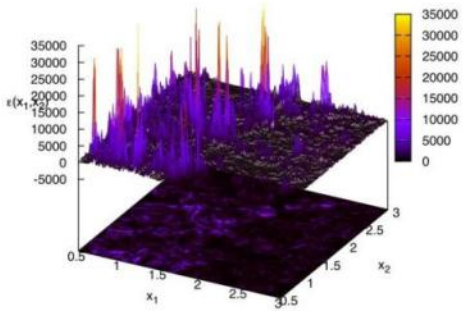


Immersive Turbulence

"… the last unsolved problem of classical physics…" Feynman

- **Understand the nature of turbulence**
  - *Consecutive snapshots of a large simulation of turbulence: now 30 Terabytes*
  - *Treat it as an experiment, **play** with the database!*
  - ***Shoot test particles** (sensors) from your laptop into the simulation, like in the movie Twister*
  - *Next: 70TB MHD simulation*





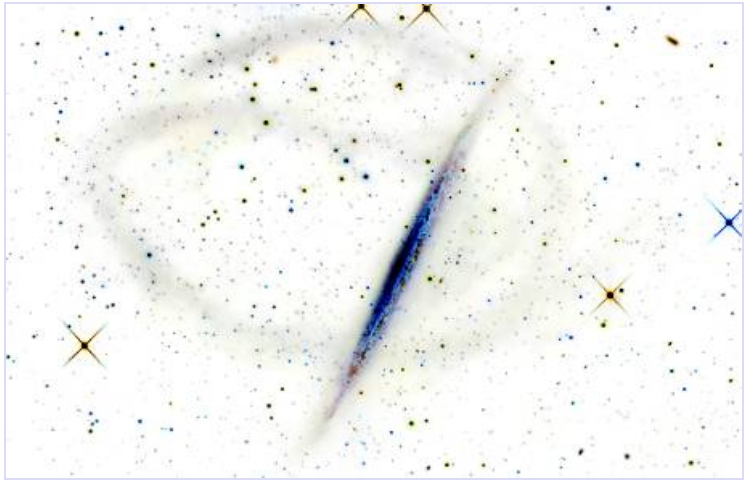
- **New paradigm** for analyzing simulations!

with C. Meneveau, S. Chen (Mech. E), G. Eyink (Applied Math), R. Burns (CS)

The Milky Way Laboratory

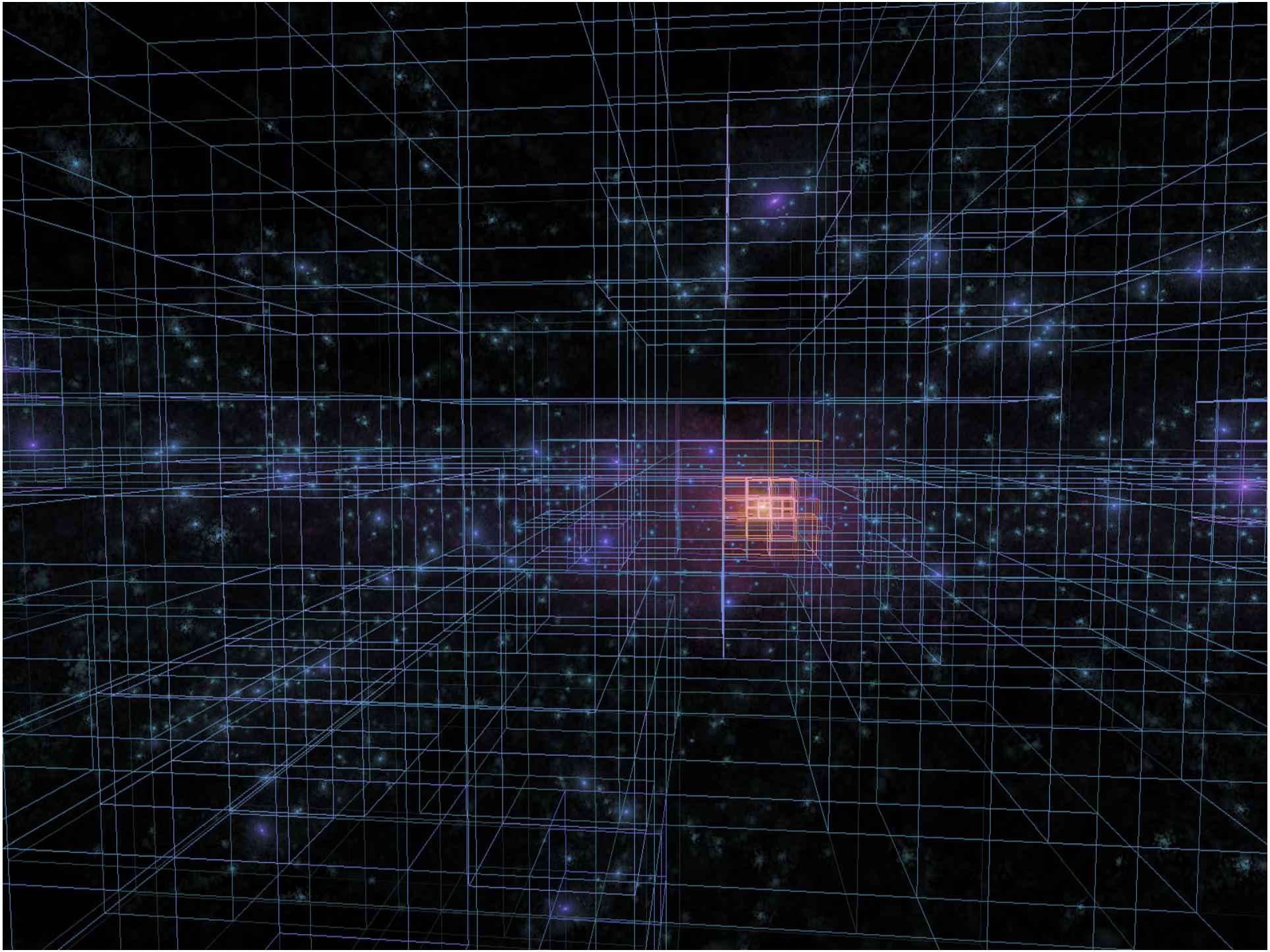
- Use cosmology simulations as an immersive laboratory for general users
- Via Lactea-II (20TB) as prototype, then Silver River (50B particles) as production (15M CPU hours)
- 800+ hi-rez snapshots (2.6PB) => 800TB in DB
- Users can insert test particles (dwarf galaxies) into system and follow trajectories in pre-computed simulation
- Users interact remotely with a PB in 'real time'

Madau, Rockosi, Szalay, Wyse, Silk, Lemson, Westermann, Blakeley

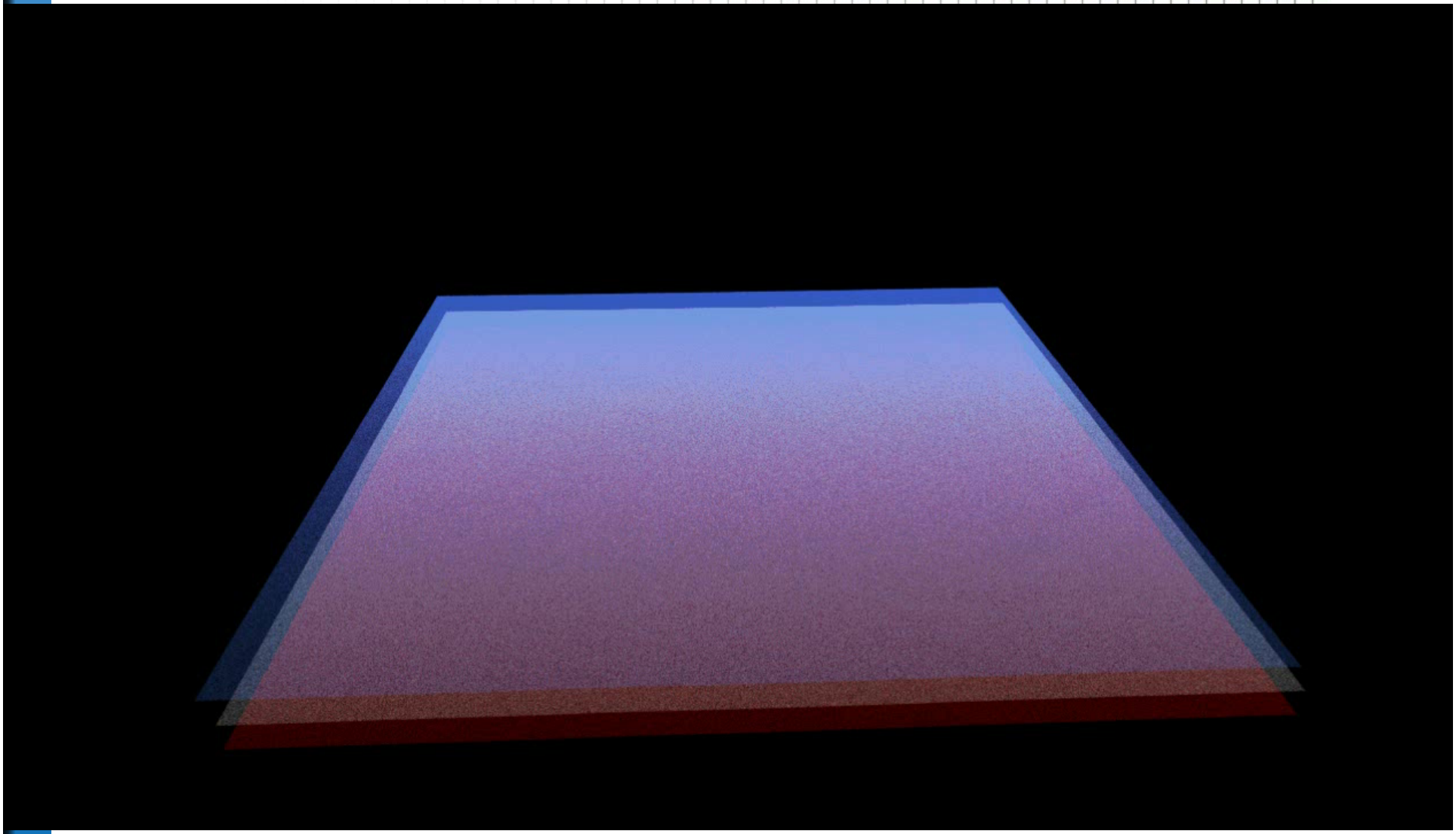


Visualizing Petabytes

- Needs to be done where the data is…
- It is easier to send a HD 3D video stream to the user than all the data
- Interactive visualizations driven remotely
- Visualizations are becoming IO limited: precompute octree and prefetch to SSDs
- It is possible to build individual servers with extreme data rates (5GBps per server… see Data-Scope)
- Prototype on turbulence simulation already works: data streaming directly from DB to GPU
- N-body simulations next



3D Vorticity in a Turbulent Flow



Kai Buerger, Technische Universitat Munich

Amdahl's Laws

Gene Amdahl (1965): Laws for a balanced system

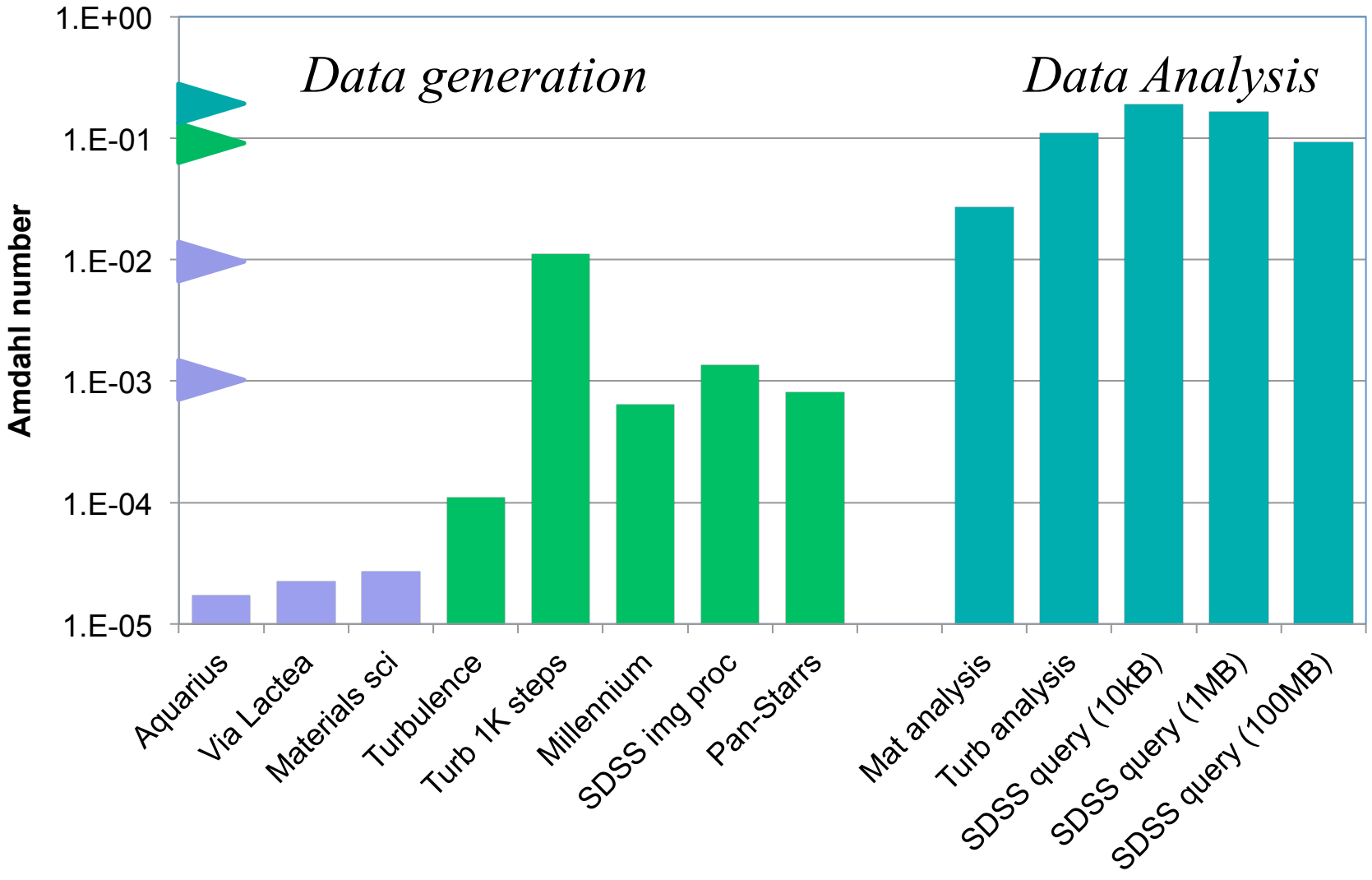
i. Parallelism: max speedup is S/(S+P)

ii. One bit of IO/sec per instruction/sec (BW)

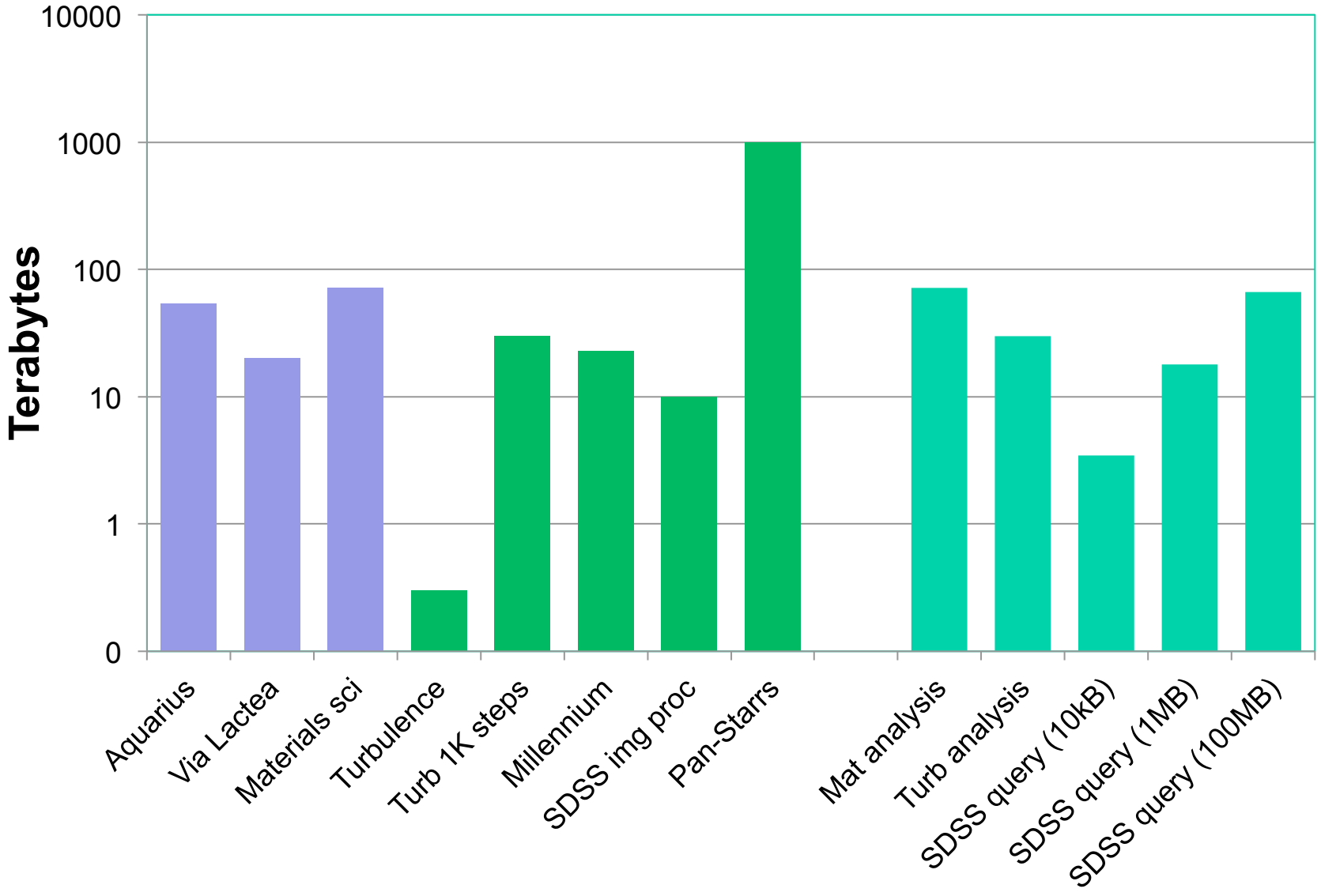
iii. One byte of memory per one instruction/sec (MEM)

Modern multi-core systems move farther away from Amdahl's Laws (Bell, Gray and Szalay 2006)

Amdahl Numbers for Data Sets



The Data Sizes Involved



DISC Needs Today

- Disk space, disk space, disk space!!!!
- Current problems not on Google scale yet:
  - *10-30TB easy, 100TB doable, 300TB really hard*
  - *For detailed analysis we need to park data for several months*
- Sequential IO bandwidth
  - *If not sequential for large data set, we cannot do it*
- How do can move 100TB within a University?
  - *1Gbps 10 days*
  - *10 Gbps 1 day (but need to share backbone)*
  - *100 lbs box few hours*
- From outside?
  - *Dedicated 10Gbps or FedEx*

Silver River Transfer

- 150TB in less than 10 days from Oak Ridge to JHU using a dedicated 10G connection

Tradeoffs Today

"Extreme computing is about tradeoffs"

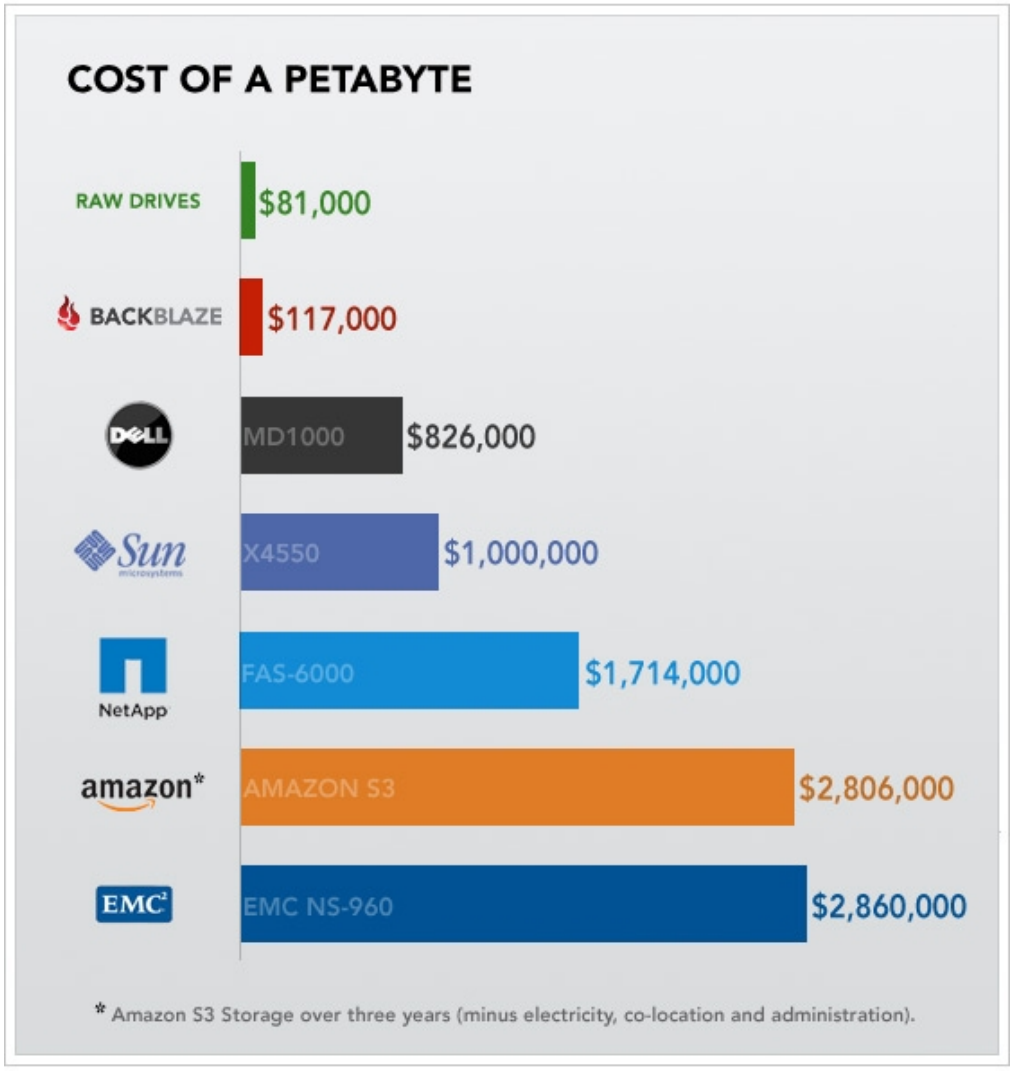
Stu Feldman (Google)

Ordered priorities for data-intensive scientific computing

1. *Total storage (-> low redundancy)*
2. *Cost (-> total cost vs price of raw disks)*
3. *Sequential IO (-> locally attached disks, fast ctrl)*
4. *Fast stream processing (->GPUs inside server)*
5. *Low power (-> slow normal CPUs, lots of disks/mobo)*

The order will be different in a few years...and scalability may appear as well

Cost of a Petabyte



From backblaze.com
Aug 2009

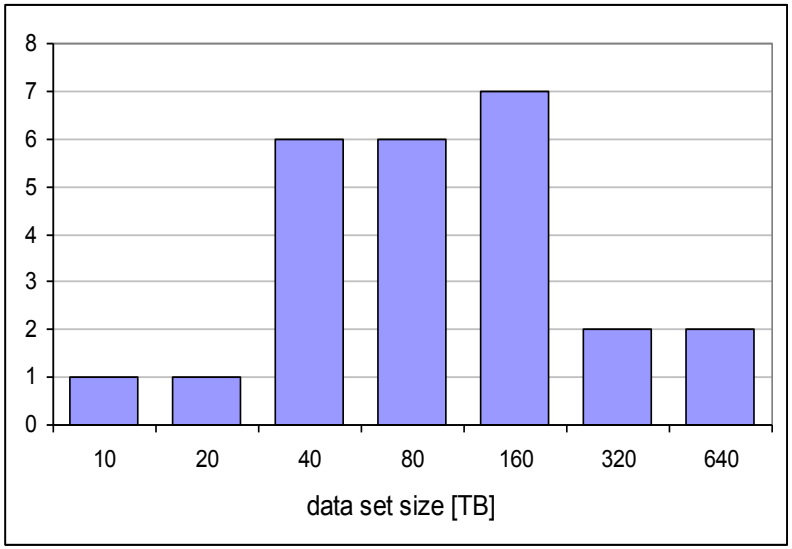
JHU Data-Scope

- Funded by NSF MRI to build a new 'instrument' to look at data
- Goal: 102 servers for $1M + about $200K switches+racks
- Two-tier: performance (P) and storage (S)
- Large (5PB) + cheap + fast (400+GBps), but …
  . ..a special purpose instrument

	1P	*1S*	*90P*	*12S*	*Full*	
servers	1	1	90	12	102	
rack units	4	12	360	144	504	
capacity	24	252	2160	3024	5184	TB
price	8.5	22.8	766	274	1040	$K
power	1	1.9	94	23	116	kW
GPU	3	0	270	0	270	TF
seq IO	4.6	3.8	414	45	459	GBps
netwk bw	10	20	900	240	1140	Gbps

Proposed Projects at JHU

Discipline	data [TB]
Astrophysics	930
HEP/Material Sci.	394
CFD	425
BioInformatics	414
Environmental	660
Total	2823



19 projects total proposed for the Data-Scope, more coming, data lifetimes between 3 mo and 3 yrs

Increased Diversification

One shoe does not fit all!

- Diversity grows naturally, no matter what
- Evolutionary pressures help
  - *Large floating point calculations move to GPUs*
  - *Large data moves into the cloud (private or public)*
  - *RandomIO moves to Solid State Disks*
  - *Stream processing emerging (SKA…)*
  - *noSQL vs databases vs column store vs SciDB …*
- Individual groups want subtle specializations

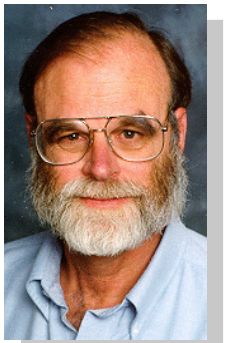
At the same time

- What remains in the middle (common denominator)?
- Boutique systems dead, commodity rules
- We are still building our own…

Summary

- Science is increasingly driven by large data sets
- Large data sets are here, COTS solutions are not
  - *100TB is the current practical limit*
- We need a new instrument: a “microscope” and “telescope” for data=> a **Data-Scope**!
- Increasing diversification over commodity HW
- Changing sociology:
  - *Data collection in large collaborations (VO)*
  - *Analysis done on the archived data, possible (and attractive) for individuals*
- A new, Fourth Paradigm of Science is emerging…

but it is not incremental….



"*If I had asked my customers what they wanted, they would have said faster horses…*"

Henry Ford

From a recent book by Eric Haseltine:
"Long Fuse and Big Bang"